

**Structural classification of *Quillaja*
saponins by electrospray ionization ion
trap multiple-stage mass spectrometry
in combination with multivariate
analysis**

Johan Bankefors

*Department of Chemistry, SLU
Uppsala*

**Licentiate thesis
Swedish University of Agricultural Sciences
Uppsala 2006**

ISBN 91-576-7181-8
© 2006 Johan Bankefors, Uppsala
Tryck: SLU Service/Repro, Uppsala 2006

Abstract

Bankefors, J. 2006. Structural classification of *Quillaja* saponins by electrospray ionisation ion trap multiple-stage mass spectrometry in combination with multivariate analysis. Licentiate thesis.

ISBN 91-576-7181-8.

This thesis describes methods for structural classification of *Quillaja* saponins with electrospray ionisation ion trap multiple-stage mass spectrometry, in combination with multivariate analysis. The mass spectrometry method was optimised by the use of design of experiments.

47 of previously reported *Quillaja* saponins from the chromatographic fractions QH-A, QH-B, and QH-C have been investigated. MS¹-MS³ spectra were analysed by multivariate methods such as PCA and PLS-DA. Fragmentation of saponins generally results in loss of end elements from the precursor ion. The essential part of this method is the re-referencing of spectra. Peaks in the obtained re-referenced spectra have a correlation to loss of common structural elements. The multivariate methods captured the variance corresponding to the common structural elements. Thus, the obtained models have the ability to predict new compounds that share the common structural elements.

Two saponins previously not characterised were isolated and analysed by ESI-IT-MSⁿ, the multivariate models predicted the structure. The obtained models were applied to HPLC on-line coupled MSⁿ data and a rapid method for screening of saponin fractions was developed.

Keywords: electrospray ionization ion trap multiple-stage mass spectrometry, *Quillaja* saponins, oligosaccharide sequencing, structural determination, PCA, PLS-DA.

Authors address: Johan Bankefors, Department of chemistry, SLU, PO Box 7015, SE-750 07 Uppsala, Sweden. E-mail:Johan.Bankefors@kemi.slu.se

Table of Contents

General introduction	9
<i>Saponins</i>	
<i>Quillaja</i> Saponins as adjuvants with vaccines	10
Structural determination of saponins	10
<i>Reported structures classified in this thesis</i>	
<i>Separation of saponins</i>	
<i>Structural characterisation of saponins</i>	
<i>Mass spectrometry on saponins</i>	
<i>Design of experiment, DOE</i>	
<i>Spectral data obtained</i>	
<i>Multivariate analysis of mass spectra</i>	
<i>Introduction</i>	
<i>Classification of saponins by multivariate modelling</i>	
<i>Differentiation of apiose from xylose as R^4-substituent.</i>	
<i>Classification of previously unreported saponins</i>	
<i>LC/MS on chromatographic fractions containing saponins</i>	
Concluding remarks	40
References	41
Acknowledgments	44

Appendix

Papers I and II

This thesis is based on the following papers, which will be referred to by their Roman numerals:

- I.** “Structural classification of *Quillaja* saponins by electrospray ionization ion trap multiple-stage mass spectrometry in combination with multivariate analysis”

Johan Bankefors, Lars I. Nord, Lennart Kenne. Submitted to Journal of Mass Spectrometry.

- II.** “Structural classification of fatty acyl substituted *Quillaja* saponins by electrospray ionization ion trap multiple-stage mass spectrometry in combination with multivariate analysis”

Johan Bankefors, Lars I. Nord, Lennart Kenne. In manuscript form.

Abbreviations

^1H	Proton
Ac	Acetyl
Api	Apiose
Ara	Arabinose
CCD	Central composite design
CID	Collision-induced dissociation
D_i	Discriminatory power
Da	Dalton
DOE	Design of experiment
Fa	Fatty acyl
ESI	Electrospray ionisation
Fuc	Fucose
Gal	Galactose
GC	Gas chromatography
Glc	Glucose
GlcA	Glucuronic acid
ISCOM	Immunonstimulating complex
IT	Ion trap
LOO-CV	Leave-one-out cross-validation
LV	Latent variable
MALDI-TOF	Matrix assisted laser desorption/ionisation-time of flight
M_i	Modelling power
MS	Mass spectrometry
MS ⁿ	Multiple-stage mass spectrometry
m/z	Mass-to-charge-ratio
NMR	Nuclear magnetic resonance
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares projections to latent structures
PLS-DA	Partial least squares projections to latent structures-discriminant analysis
PRESS	Predicted residual error sum of squares
Q_{cv}^2	Squared correlation coefficient for cross-validation
$Q_{test-set}^2$	Squared correlation coefficient for test set predictions
R^2	Squared correlation coefficient for calibration set fitting
Rha	Rhamnose
RP-HPLC	Reversed phase high performance liquid chromatography
S_{pooled}	Pooled estimate of standard deviation
SIMCA	Soft independent modelling of class analogy
SNV	Standard normal variate correction
UV	Ultra violet
Xyl	Xylose

General introduction

The scope of this licentiate thesis is to demonstrate the usefulness of multivariate methods in combination with mass spectrometry to obtain a method for rapid structural classification. The aim is to provide a method that is directly applicable to *Quillaja* saponins; however, this approach should also be relevant to other substances built of different monomers or repeating units.

Saponins

Saponins are a group of steroidal and triterpene glycosides that are widely found in plants, one commonly known is the ginseng plant that has been in use partly due to its haemolytic effects. However, there exist only a few plants that provide a larger source of saponins, hence are worth extracting for commercial use. The *Quillaja saponaria* Molina tree, which is native to Chile, Peru and Bolivia, is one of these sources. The bark contain of up to 5% by weight of saponins that makes it one of the strongest commercial sources. The widespread use of saponins extracted from this bark has led to a discussion about how to develop a sustainable exploration [1]. The applications of saponins stretch over several areas such as additives in food and cosmetics, as wetting agents for the agriculture and photographic industry and as adjuvants in the pharmaceutical industry [1].

Extracts of bark contain a complex mixture of triterpenoid saponins that consist of a quillaic acid as the aglycone, which is substituted with a di- or trisaccharide at C-3 (X^1) and an oligosaccharide at C-28 with an X^2 -, X^3 -, and X^4 -substituted fucose as the first monomer (Fig. 1).

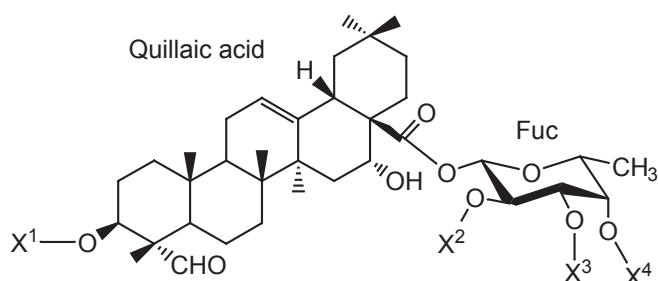


Fig. 01. Common basic structure reported for *Quillaja* saponins.

***Quillaja* Saponins as adjuvants with vaccines**

The bark of *Quillaja saponaria* Molina has been shown to contain the most potent adjuvants of several investigated plants [2]. Several methods for enrichment have been carried out using techniques such as dialysis [2], gel permeation [3] or diafiltration [4]. However, the fractions obtained this way are heterogeneous and their biological and chemical activities are not very predictable. This might be due to the fact that 20-25% of the extractable material in bark varies in content of saponins [5]. The first purified saponin fraction with predictable adjuvant properties was termed Quil A and the separation was achieved with anion exchange chromatography and gel filtration [6]. Quil A had in contrast to earlier fractions a more predictable adjuvant activity and was less toxic in veterinary applications. However, the use of a heterogeneous mixture such as Quil A is not satisfactory on humans since the toxicity and adjuvant activity still vary with the existence and concentration of different saponins. Several methods for separation and identification of saponin classes in Quil A were carried out utilizing RP-HPLC [4, 7-9]. Rönnberg *et al.* separated Quil A into three fractions using RP-HPLC which were designated QH-A, QH-B and QH-C due to elution order. The fractions were tested and were found to have different adjuvant activity and toxicity with QH-B as the more toxic fraction whilst QH-C is the one with most potent adjuvant activity. This activity has later been linked to the fatty acyl substituent on the fucose residue. The fatty acyl substituent has a large impact on the adjuvant activity and toxicity [10].

Usually the best effect of a vaccine is reached by using live, attenuated or killed microorganisms, as an alternative approach a purified antigen can be used from a virus, bacteria or parasite. The latter approach is considered to be safer but induce less immune response. Therefore, adjuvants can be used to compensate for this loss of response. There is a wide range of adjuvant systems starting with aluminium salts that were reported in 1925 [11]. Veterinary applications of *Quillaja* saponins have been included in several of these. During the last decade a new system called ISCOM[®]s (Immunostimulating complex) was formulated by Morein *et al.* [12]. ISCOM[®]s are composed of antigen, cholesterol, phospholipids and saponins and a variant of ISCOM[®] without the antigen is called ISCOMATRIX[®]. ISCOM[®]s have shown to induce an antibody and cellular immune response in animals [13, 14] and show promising results for being developed into an effective human vaccine [15].

The adjuvant component of ISCOM[®]s is a mixture of saponins from chromatographic fractions of QH-A and QH-C avoiding the more toxic QH-B fraction. ISCOM[®]s with different *Quillaja* saponin components differ in their immunomodulating activities and toxicities, hence it is of interest to study the composition of saponins and separate and determine their structures. This can however be a very time consuming task and thus rapid and sensitive methods for screening of *Quillaja* saponins are important for further understanding of its effects.

Structural determination of saponins

Reported structures classified in this thesis

To date, the complete structures of 67 native *Quillaja* saponaria Molina have been reported [16-23], 47 of these saponins were studied in this work. The common basic structure is illustrated in Fig 1. All studied structures consist of the quillaic acid that is substituted with a di- or tri-saccharide at C-3 and a branched oligosaccharide at C-28 (outlined in Fig. 2 and Fig. 3). Glucuronic acid, substituted at its C-2 with galactose, is linked to C-3 of the quillaic acid. The glucuronic acid is further substituted at its C-3 with either a rhamnose or a xylose. The R¹ substituent is the reason why some saponins elute in pairs with a difference of 14 mass units. The common structure of the oligosaccharide at C-28 consists of a fucose O-2 substituted by a rhamnose at C-2 and the latter is further substituted by a xylose at C-4. The fucose is O-4 acylated, either with an acetyl or a fatty acyl group. R substituents are numbered in order to be comparable to paper I (Fig. 2) and paper II (Fig. 3). The saponins included in this thesis are most of the major components in the chromatographic fractions QH-A, QH-B and QH-C. Components in QH-A were investigated in paper I, and those in QH-B and QH-C in paper II. The numbering of the compounds is preserved from the original publications.

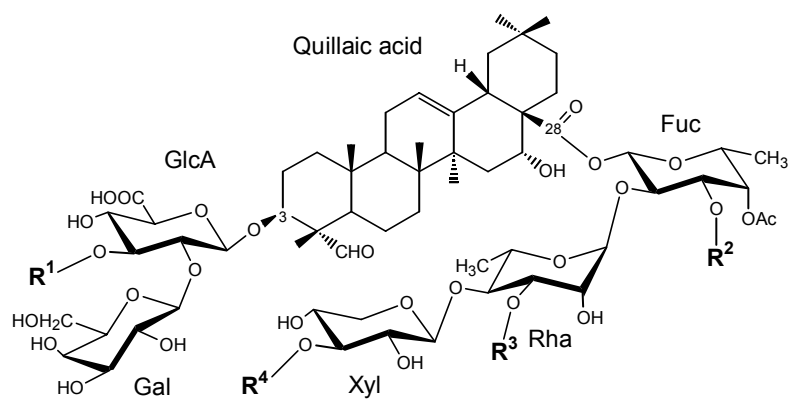


Figure 02. Basic structure of the investigated *Quillaja* saponins in QH-A. Substituents in R^1 to R^4 are shown in Table 1.

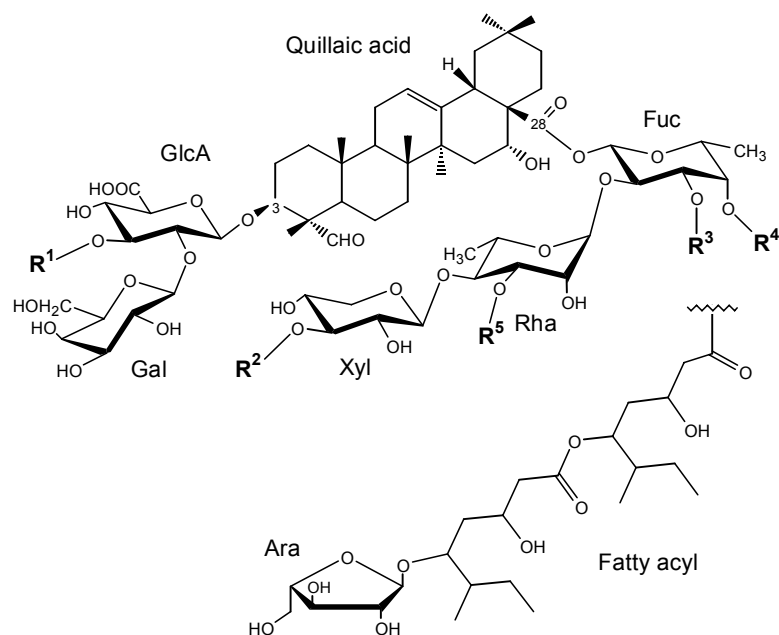


Figure 03. Basic structure of the investigated *Quillaja* saponins in QH-B and QH-C. Substituents in R^1 to R^5 are shown in Table 2.

Table 01. *Substituents of the investigated saponins in QH-A (Fig. 2)*

Compound	R ¹	R ²	R ³	R ⁴	Mass ^a
4	H	Rha	H	H	1436.7
5	Rha	Rha	H	H	1582.7
6	Xyl	Rha	H	H	1568.7
7	Rha	Rha	H	Api	1714.8
8	Xyl	Rha	H	Api	1700.7
9	Rha	Rha	H	Xyl	1714.8
10	Xyl	Rha	H	Xyl	1700.7
11a,b	Rha, Xyl	Glc	H	H	1598.7, 1584.7
17a,b	Rha, Xyl	Rha	Glc	H	1744.8, 1730.7
18a,b	Rha, Xyl	Rha	Glc	Api	1876.8, 1862.8

^a Reported monoisotopic molecular massesTable 02. *Substituents of the investigated saponins in QH-B and QH-C, (Fig. 3)*

Compound	R ¹	R ²	R ³	R ⁴	R ⁵	Mass ^a
B1	Rha	H	H	Fatty acyl	Glc	2033.0
B1a	Rha	H	Fatty acyl	H	Glc	2033.0
B2	Xyl	H	H	Fatty acyl	Glc	2018.9
B2a	Xyl	H	Fatty acyl	H	Glc	2018.9
B3	Rha	Api	H	Fatty acyl	Glc	2165.0
B3a	Rha	Api	Fatty acyl	H	Glc	2165.0
B4	Xyl	Api	H	Fatty acyl	Glc	2151.0
B4a	Xyl	Api	Fatty acyl	H	Glc	2151.0
B5	Rha	Xyl	H	Fatty acyl	Glc	2165.0
B5a	Rha	Xyl	Fatty acyl	H	Glc	2165.0
B6	Xyl	Xyl	H	Fatty acyl	Glc	2151.0
B6a	Xyl	Xyl	Fatty acyl	H	Glc	2151.0
B7	H	H	H	Fatty acyl	Glc	1886.9
B7a	H	H	Fatty acyl	H	Glc	1886.9
B8	H	Api	H	Fatty acyl	Glc	2018.9
B8a	H	Api	Fatty acyl	H	Glc	2018.9
S1	Rha	H	H	Fatty acyl	H	1870.9
S1a	Rha	H	Fatty acyl	H	H	1870.9
S2	Xyl	H	H	Fatty acyl	H	1856.9
S2a	Xyl	H	Fatty acyl	H	H	1856.9
S3	Rha	Xyl	H	Fatty acyl	H	2002.9
S3a	Rha	Xyl	Fatty acyl	H	H	2002.9
S4	Xyl	Xyl	H	Fatty acyl	H	1988.9
S4a	Xyl	Xyl	Fatty acyl	H	H	1988.9
S5	Rha	Api	H	Fatty acyl	H	2002.9
S5a	Rha	Api	Fatty acyl	H	H	2002.9
S6	Xyl	Api	H	Fatty acyl	H	1988.9
S6a	Xyl	Api	Fatty acyl	H	H	1988.9
S7	Rha	H	Acetyl	Fatty acyl	H	1912.9
S8	Xyl	H	Acetyl	Fatty acyl	H	1898.9
S9	Rha	Xyl	Acetyl	Fatty acyl	H	2045.0
S10	Xyl	Xyl	Acetyl	Fatty acyl	H	2030.9
S11	Rha	Api	Acetyl	Fatty acyl	H	2045.0
S12	Xyl	Api	Acetyl	Fatty acyl	H	2030.9

^a Reported monoisotopic molecular masses

Separation of saponins

With time, several analytical methods have been used for both structural elucidation and quality control purposes. RP-HPLC with UV detection[4] is commonly used for separation, a drawback using this method is low sensitivity because saponins lack strong chromophores. However, studies have shown that an RP-HPLC system meets the requirements for most separations of *Quillaja* saponins. The retention time for saponins is sensitive to concentration of organic modifier and pH. The isolation of saponins from QH-A, QH-B, and QH-C is normally a two-step procedure. Initially the saponins are separated due to structural features in the C-28 oligosaccharide at pH 5.8-6.5. In the second step at pH 2.8, the co-eluted saponins with differences in the C-3 oligosaccharide are separated. This behaviour might be due to the glucuronic acid in the C-3 di- or trisaccharide. Fig. 4 visualises the two chromatographic steps of QH-C and QH-B. The *Quillaja* saponin bark extracts used in this study were obtained from commercial fractions QH-A, QH-B and QH-C (Iscotec, Uppsala). Saponins in QH-A were previously isolated by Guo *et al.*[18, 19] A preparative and semi-preparative scale HPLC separated QH-B and QH-C. Due to migration of the fatty acyl substituent mass spectrometric analysis followed the final separation step of the fractions QH-B and QH-C without delay to ensure that the sample contained only one of the regio-isomers. The phosphate buffer systems previously used to separate the region-isomers at low pH (2.8)[21, 22] were changed to a more volatile formate buffer to be in compliance with the liquid chromatography mass spectrometry method later used.

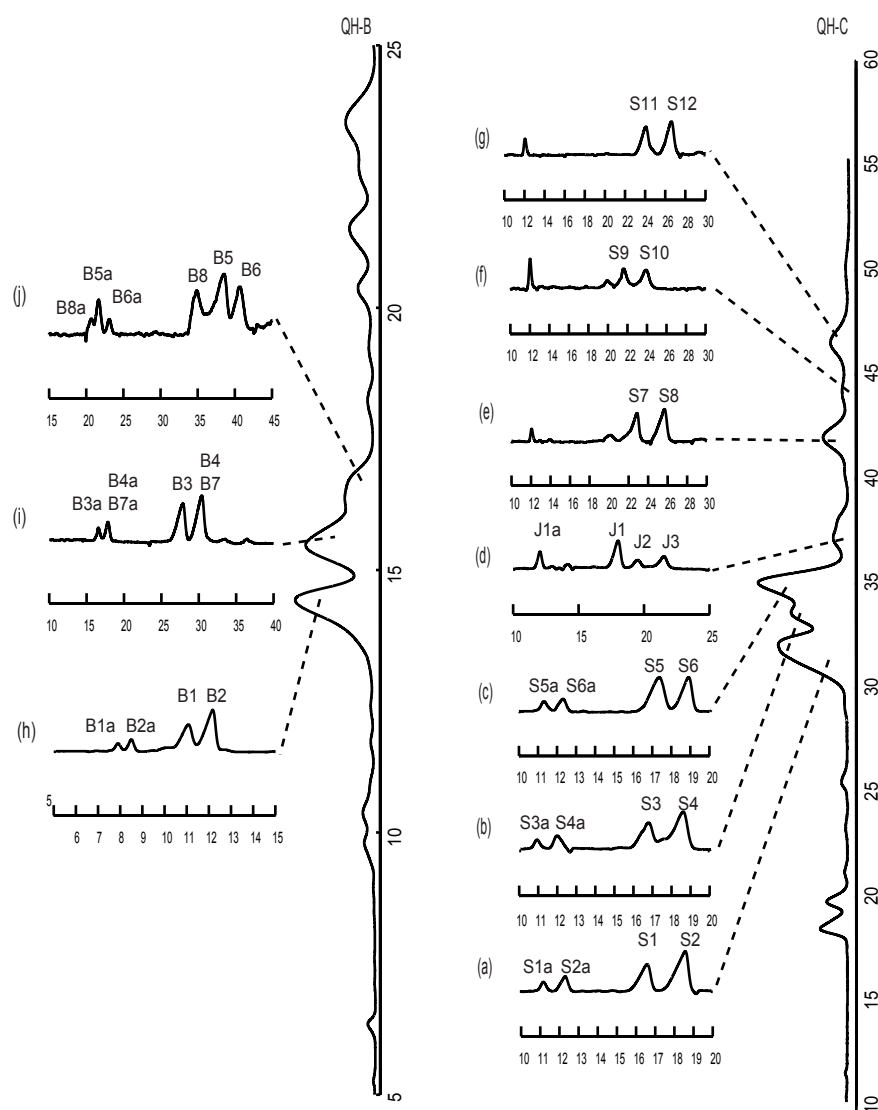


Fig. 04. HPLC chromatogram illustrating the two chromatographic steps of saponin fractions QH-C and QH-B with sub-fractions a-j. First separation step was carried out with a phosphate buffer at pH 6.4, in the second step a formate buffer at pH 2.8 was used. **J1a, J1, J2, J3** are new compounds visualised in (d), **J1a** and **J1** were positively identified by the method described. **J2** and **J3** were dissimilar thus not recognised by the models.

Structural characterisation of saponins

So far, NMR techniques have mainly been used to achieve a full elucidation of the reported structures of saponins. This can however be a time consuming task and the phenomenon of overlapping signals usually troubles the investigation of saponins with NMR. Commonly hydrolysis, and derivatization[24] followed by GC-MS [25] and nanoelectrospray ion trap multiple-stage mass spectrometry [26, 27] and have been used as a complementary pre-NMR tool. Nord *et al.* [28] used multivariate analysis in combination with ^1H NMR spectra to achieve a method for classification of saponins.

Another approach of structural investigations is based on the fact that *Quillaja* saponins are combined of certain common structural elements. Hence it is possible to classify structures by mass spectrometry, without a full scale NMR characterisation procedure.

The possibilities of using ESI-MSⁿ on a mixture of triterpenoid saponins extracted from leaves of *Acanthopanax senticosus* Harms were investigated [29]. It was found that this procedure confirmed known structures and provided useful structural information on the components, but structures for unknowns were not proposed. Isolated compounds of *Quillaja* saponins from QH-A, QH-B and QH-C were extensively investigated using ESI-ITMSⁿ [30, 31]. Structural information was provided when fragmentation routes were elucidated and key fragments were proposed. This study provided a method to characterize the structures in a specific group of non-derivatized *Quillaja* saponins, based on MS¹-MS³ experiments in the positive mode.

Liquid chromatography coupled with on-line mass spectrometry has been applied to identify and authenticate saponins in crude extracts together with metabolomic profiling [32, 33]. Structures were investigated using HPLC-UV-MS/MS, key fragments that are characteristic to certain classes of saponins such as the fatty acyl domain and the aglycone were pinpointed to provide evidence of the structure. The approach of using LC/MS for metabolomic studies of *Quillaja* saponins is promising due to the increased resolution and sensitivity obtained with both chromatographic and mass separation.

Previous studies of *Quillaja* saponins using on-line coupled RP-HPLC-ESI-ITMSⁿ for characterisation of structures have not progressed further than MS² either in positive or negative ion mode [33-36] hence the oligosaccharide structure at C-28 has not thoroughly been investigated.

In order to combine structural identification with HPLC methods a rapid on-line method is required.

Reported mass spectrometry methods have until today been manually evaluated and thereby time consuming. By combining the HPLC-MSⁿ methods with multivariate methods the time required for a structural classification of known and unknown saponins will decrease.

In paper I a method for structural classification using multivariate methods of most of the major saponins in QH-A is outlined.

Paper II adds the LC/MS profiling with the structural characterisation of *Quillaja* saponins into one method. The LC/MS procedure can be used as a profiling step while the multivariate models obtained from known structures in both paper I and paper II have the potential of structural classification of a large part of the major known and unknown compounds in crude extracts or chromatographic fractions such as Quil A.

Mass spectrometry on saponins

In paper I and II both MALDI-TOF and ESI-ITMSⁿ have been utilized for studies on saponins with the emphasise on ESI-ITMSⁿ. The area of application is different between the two but they have a few things in common, both belong to the group of so-called soft ionisation techniques. When an analyte is subjected to soft ionisation they almost exclusively form molecular ions thus initial fragmentation is not obtained.

MALDI-TOF

There are different sample preparation methods in MALDI-TOF MS [37], but commonly the analyte is dissolved in a light-absorbing matrix. The matrix with the embedded analyte is then targeted with a UV-laser beam whereupon the energy is absorbed and gently transferred by the matrix to the analyte that is ionised. This mechanism is not fully understood but since most matrices are acids some believe that by transferring a proton to the analyte they can promote the ionisation. When the analyte has been irradiated and formed ions these are accelerated by high voltage and separated based on the time it takes them to drift through a field free region. Since all ions will receive the same kinetic energy the ions with lowest molecular weight will travel faster over the field free region and reach the detector first. To increase the mass resolution of MALDI-TOF measurements an ion mirror or reflectron can be used. This device refocuses the ions of similar m/z so that the arrival time distribution is reduced at the detector.

ESI-ITMSⁿ

ESI is an ionisation method by which the analyte is charged and transferred from a solution to the gas phase. Generally, the process involved in this formation of ions needs an electric field to the tip of a capillary containing a solution of electrolyte ions. Depending on the field a double layer will form with either positive or negative ions near the surface. When the capillary is positive, positive electrolyte ions will be near the surface and destabilise the meniscus and a cone is formed. The cone will eventually break up into small multiple charged droplets. The solvent will evaporate from droplets and the shrinkage will lead to coulombic repulsion, at a stage when the repulsion is high enough to overcome the surface tension droplet fission takes place. Eventually very small droplets will form, however the final formation of the singly charged molecular ion is not fully understood. There exist two accepted mechanisms, the charged residue model and the ion evaporation model [38-40].

The ions transferred into gas phase have to be focused by ion-optics from the needle into the ion trap. This is done by a capillary-skimmer-octopole system.

The drying gas, heated capillary and pressure gradient all contribute to removing the remaining solvent ion-clusters. The skimmers then let ions through to the octopole while removing most of the drying gas. The octopole acts as a selective filter where ions in a certain m/z range are stabilised depending on voltage. Optimizing the ion-optics conditions is important for the efficiency of ion transport as well as for the signal to noise ratio obtained and appearance of spikes in spectra.

When ions are collected in the quadrupole ion trap they can either be ejected and detected or isolated and fragmented. An additional voltage accelerates the ions. Fragmentation is achieved when collision occurs between the accelerated precursor ions and the inert gas (collision-induced dissociation, CID) [41]. The daughter ions formed from the previously isolated and fragmented precursor ion can then be detected in the second MS stage, thus ions solely originating from the precursor ion can be monitored. This procedure can be repeated until the abundance of ions is too low to be detected.

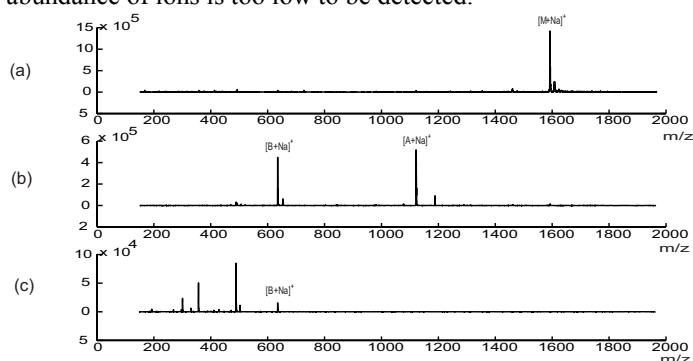


Fig. 05. MS^n spectra of compound **6** (a) MS^1 spectrum of a fraction containing compound **6** (b) MS^2 spectrum of the isolated and fragmented precursor ion $[M+Na]^+$ and the dominating ions $[A+Na]^+$ and $[B+Na]^+$. (c) MS^3 spectrum of the isolated and fragmented daughter ion $[B+Na]^+$.

The advantage with MALDI-TOF compared to ESI-ITMSⁿ is the mass accuracy and the rapid sample preparation, but on the other hand there is no possibility of fragmentation in several successive MS steps, which is the case with ESI-ITMSⁿ, which is visualised on compound **6** in Fig. 5.

All spectra reported in paper I and II were measured in positive ion mode, using ion-optic parameter settings optimised with design of experiment, DOE.

Design of experiment, DOE

Design of experiment is used in order to find out if a set of variables has an effect on one or more responses. Further, by using a systematical design understanding of a complex interdependent system can be obtained with relatively few measurements, thus providing time effectiveness and beneficial economic effect [42].

Four important aspects are implemented when applying DOE, which are replication, blocking, randomisation, and orthogonality. Replication of experiments separates the systematic variation from error of measurement. Blocking makes it possible to distinguish the impact of certain variability onto the response. Randomisation evens the impact of unknown factors over all experiments. Orthogonality between experiments allows for evaluation of independent effects [43].

Making a full investigation using DOE often involves the use of several designs depending on knowledge of the field of application. Firstly, a screening design is applied to decide the important set of factors that have a significant effect on the response. Elimination of non-significant factors that make later designs more effective. Secondly, an optimisation design is used to find best performance. The general approach is to use a design that requires as few experiments as possibly. There are numerous different designs depending on the application, the most common to start with is a full factorial design or a fractional factorial design. Since the number of experiments included in a full factorial design increases rapidly with a factor 2^k in a two level design or 3^k in a three level design, where k is the number of variables. Fractional designs are reduced full factorial designs, the advantage of using these is that less experiments have to be included. The drawback is loss of resolution, which is the ability to evaluate independent effects. In order to study curvature a three level full factorial design or a response surface design or other similar designs, which is able to adapt to and describe curvature, are required. Variables included in paper I were skimmer potential (V), octopole potential (V), trap-drive (%) and concentration of formic acid (%). The response was defined as a value produced by dividing the intensity of the $[M+Na]^+$ ion by the intensity of the ion with the second highest intensity. All measurements were outlined on compound **6**. Since the time needed for each experiment was short, a full factorial design was deployed resulting in 2^4 (16) experiments, three centroid replicates were added to the design in order to detect lack of fit and to measure the experimental error (assuming homoscedastic noise). The values of the encoded design matrix corresponds to; skimmer 50, 75, and 100 V, octopole 2, 3.5, and 5 V, trap-drive 70, 97.5, and 125%, and formic acid concentration 1, 3, and 5%. The effects of the variables on the response were evaluated and skimmer, octopole and trap-drive settings showed to be significant while the concentration of formic acid was insignificant. Fig. 6 visualises the response against the three significant variables. All three centroid replicates were obtained at a higher response than predicted by the response surface of the original full factorial design, indicating a curvature in the response.

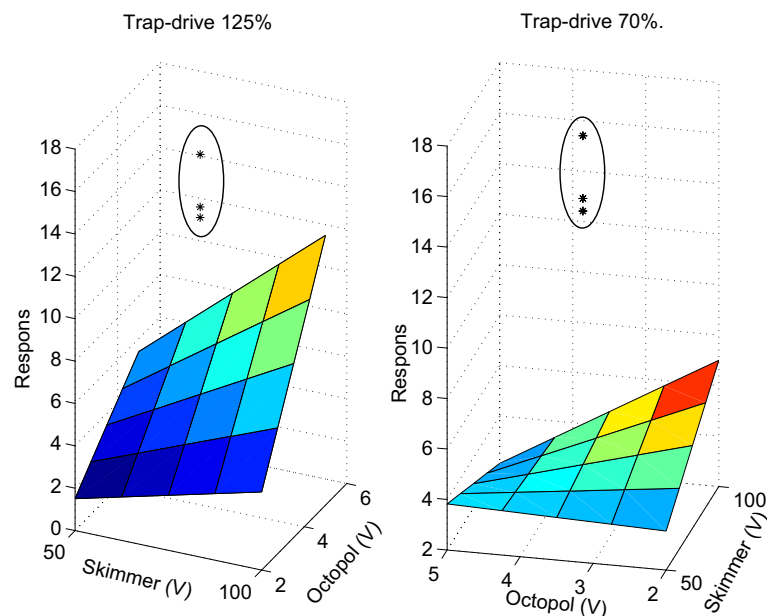


Fig. 06. Response plots of the full factorial design. The three centroids (encircled) are visualised above the response surface, indicating a response curvature.

A central composite design (CCD) was used to find the optimal parameter settings. The evaluation of the full factorial design made it possible to exclude one variable (concentration of formic acid) and detect a curvature in the response as well as adjusting the limits for the design.

The limits for the three variables were set to; skimmer 65, 70, 82.5, 95, and 100 V, octopole 2, 2.5, 3.5, 4.5, and 5 V, trap-drive 70, 78, 97.5, 117, and 125%. The effects of the variables on the response are shown in Fig. 7 and it is evident that the interaction effects have a significant effect on the performance of the method.

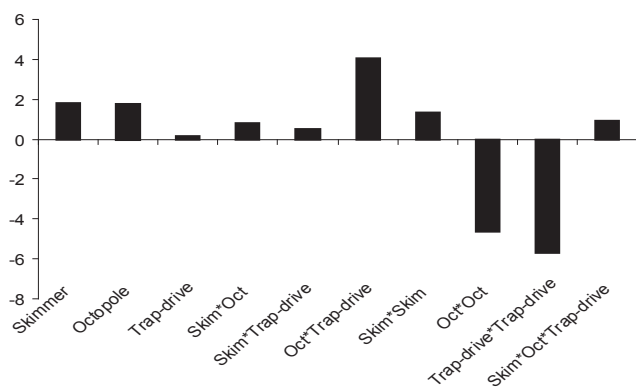


Fig. 07. *Effect of variables on the response calculated for the central composite design.*

The response surfaces (Fig. 8) were calculated, the optimum settings were found at: skimmer 100 V, octopole 4 V, trap drive 100%. These settings were then applied to all measurements in paper I and II.

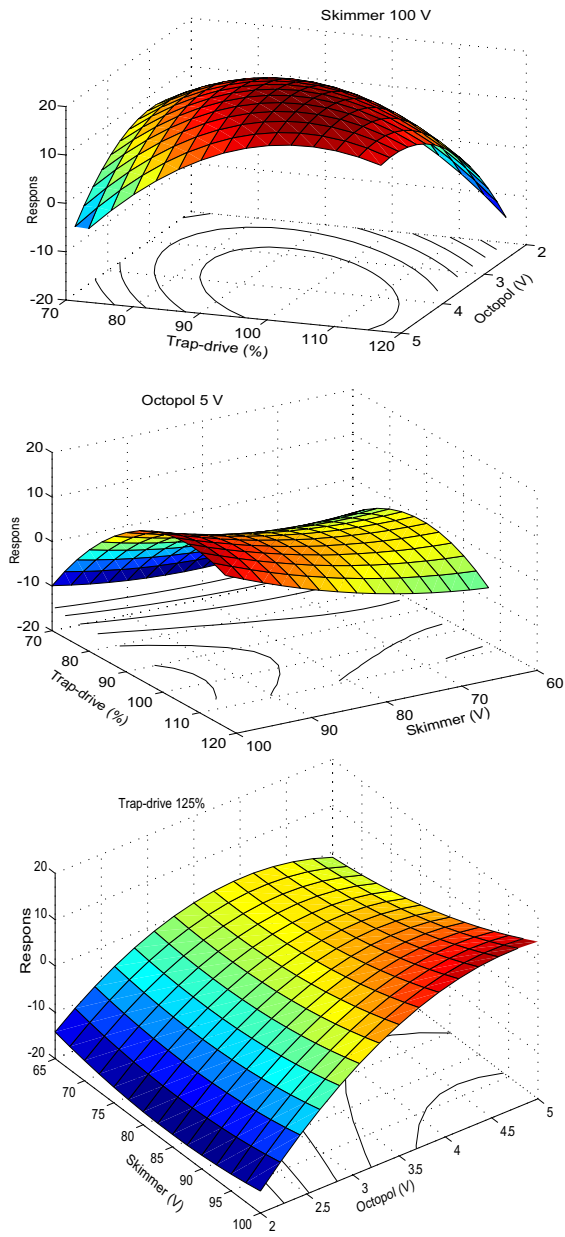


Fig. 08. Response plots of the central composite design. The optimum parameter settings were found at; skimmer 100 V, octopole 4 V, trap drive 100%.

Spectral data obtained

All compounds previously listed in Table 1 and Table 2 were investigated in positive ion mode MS^1 - MS^3 . The obtained MS^1 spectra provided the expected singly charged $[M+Na]^+$ ions. In both paper I and II isolated structures were investigated, but in paper II mixtures and chromatographic fractions were also investigated. The chromatographic and mass spectral resolution contributed to the finding of a large number of minor components in QH-B and QH-C, MS^1 spectral data vs. chromatographic resolution is visualised in Fig. 9. Each coloured area represents an eluted saponin. A distance of 132 and 162 mass units can be found between several of the saponins, corresponding to one extra pentose and hexose, respectively. The area marked with a rectangle in Fig 9(b) is further investigated in Fig 10(a) where all spectra of the region are presented and in Fig 10(b) where spectrum number 32 is selected. The saponins in QH-B and QH-C elute pair-wise, separated by 14 mass units within the pair, and 22 mass units between the pairs.

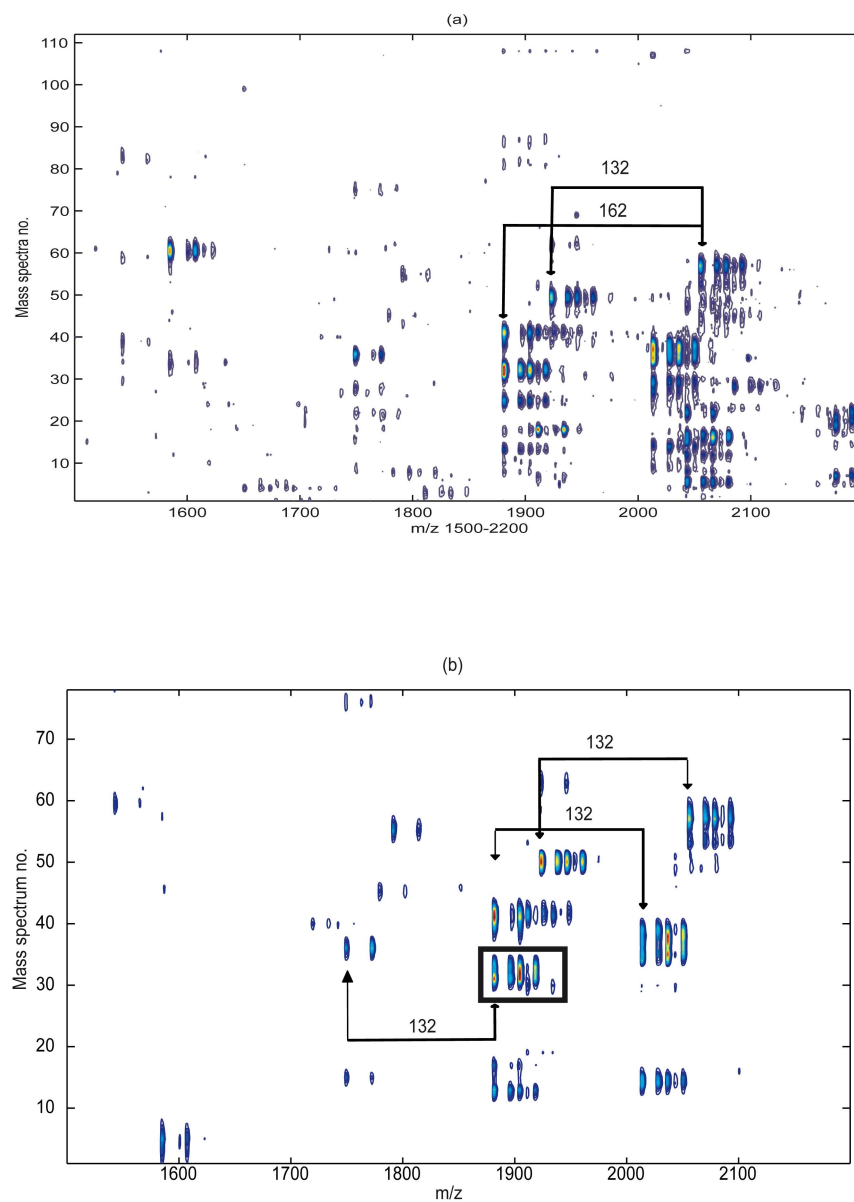


Fig. 09. Ion map obtained from HPLC on-line coupled MS^I on chromatographic fraction of QH-B (a) and QH-C (b). The selected part in (b) is further shown in Fig 10.

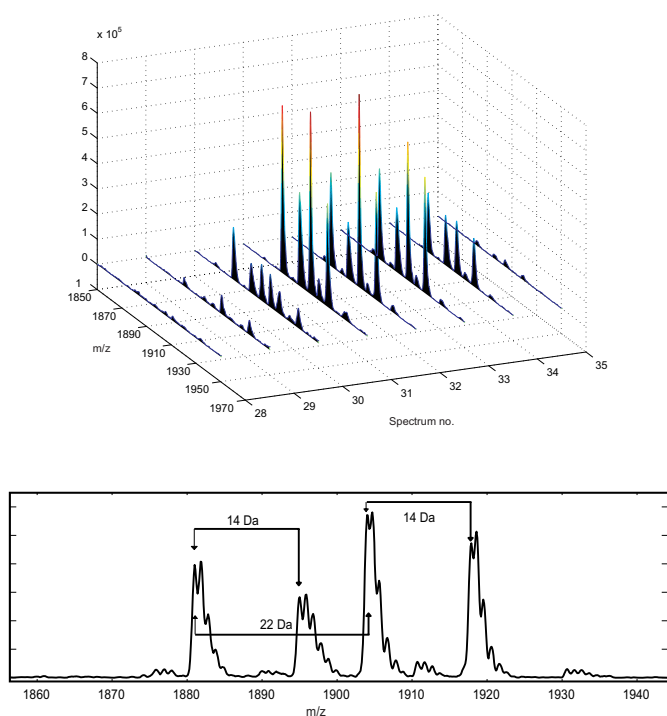


Fig. 10. Extracted spectra from Fig. 09, spectrum no. 32 is selected and visualised. Difference of 14 mass units is due to the co-eluting pairs of saponins with either xylose or rhamnose as R^1 -substituent. The difference of 22 mass units is probably due to extra sodium attached to the carboxylate in the glucuronic acid ($[M-H+2Na]^+$).

The full $[M+Na]^+$ isotopic peak pattern was isolated and fragmented in MS^2 providing the $[A+Na]^+$ and the $[B+Na]^+$ fragment ions. $[A+Na]^+$ reflects the loss of the C-3 di- or tri-saccharide whereas the $[B+Na]^+$ reflects cleavage of the bond between the C-28 oligosaccharide and the aglycone, thus both the aglycone and the di- or tri-saccharide can be classified out of MS^2 spectra.

The $[B+Na]^+$ ion was then selected, fragmented and detected in MS^3 . Theoretically, fragmentation of an ion such as the $[B+Na]^+$ leads to simultaneous cleavage of all glycoside bonds, which produce all possible daughter ions. Nonetheless, not all fragments are observed in the spectra, due to different stability of the glycoside bond and that of the formed ions. Ions formed in MS^3 can then generally be considered to belong to one of either two fragmentation routes. One route is based on a charged fucosyl residue ion, where the end elements are cleaved off while the charge remains on the fucose unit. Each loss in such a successive series will then reflect (mirror) the mass of a certain monomer. The fragmentation pattern with a perspective from the m/z of the precursor ion makes it possible to deduce the monomer sequence.

The second main route is where charged end elements are separated from the fucosyl containing oligosaccharide. Fragments of this route will always appear at the same m/z regardless of the original monomer sequence. Thus revealing elements of the structure that can be puzzled together to obtain a structure. The drawback in monitoring fragments from the second route is low sensitivity; the larger fragments containing the fucose unit have a higher abundance due to a higher stability.

Compounds included in paper I and II showed that loss of structural elements from the ends dominated [30]. The ions derived from loss of the same structural element are thus depending on the molecular mass of the original ion and are for this reason positioned at different m/z -values in the spectra. By studying a fragmentation map (Fig. 11) of observed ions it is evident that most of the detected fragments contain the fucose residue.

Fragmentation will not be further described here, for a detailed study the Ph. D. thesis by Broberg S. can be recommended.

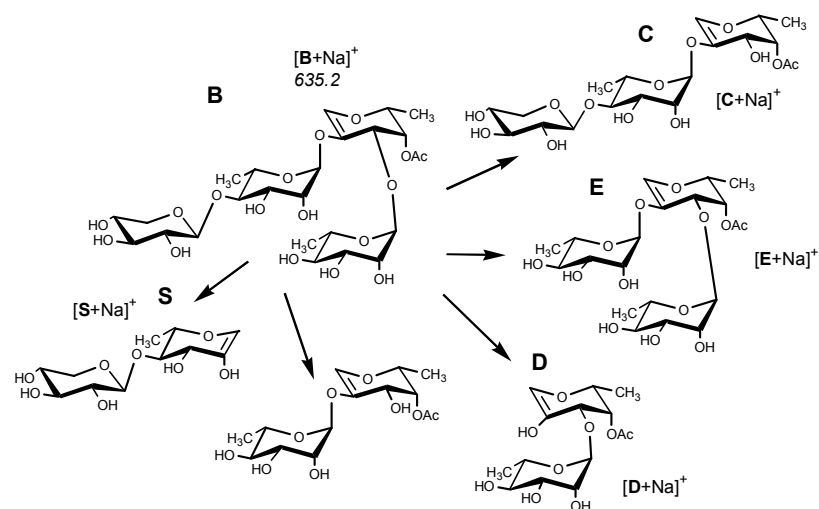


Fig. 11. Proposed fragmentation for compound **6** in MS^3 . The theoretical mass of the precursor ion $[B+Na]^+$ is given in italics. The labelling of the fragments is preserved from the original publications [30]

Multivariate analysis of mass spectra

Introduction

The pattern recognition technique is a branch of what is commonly called chemometrics. Pattern recognition has frequently been applied to various types of data where the aim is to distinguish between two or more classes. The pattern usually provides information about the relations between objects (compounds) in one class. These relations are more or less similar within one class and dissimilar between classes.

Pre-treatment of data is often required before the actual multivariate analysis in order to magnify or weaken effects contributing to the models. There are numerous techniques divided in three main areas: filters, signal correction techniques and scaling [44]. Pre-treatment in this work has been used to remove differences in intensity due to concentration or ion suppression. Pattern recognition is typically carried out in two steps, firstly a principal component analysis (PCA) [44-46] for exploring data and secondly a projection to latent structure (PLS) [47].

PCA is a data compression technique that finds orthogonal principal components (PCs), which describes major trends in the data. These PCs are ordered according to the amount of variation described so that the first PC describes the largest variation. The scores represent the objects (samples) in the new coordinate system defined by the PCs, and plotting the scores for the first few PCs against each other may discover major trends or clusters in the data. The loadings describe how the PCs are obtained from the original variables. By examination of the loadings it is possible to analyse which variables are likely to be related to observed clusters in the score plot.

A geometrical interpretation of a PC is visualised in Fig. 12. The direction of the line is described by the loading and the scores are illustrated as the open circles, which is the orthogonally projected data (closed circles are the obtained measurements).

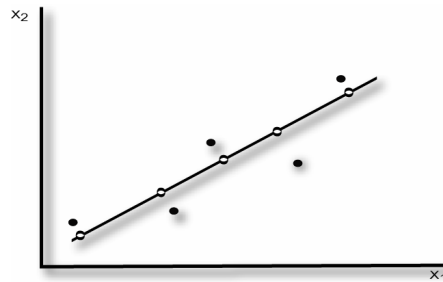


Fig 12. *A geometrical interpretation of a PC, the open circles are the projected experimental data (closed circles) onto the PC.*

When the first PC has been fitted to the data (X -matrix) the captured variation in the first PC (t_1 and p_1) is subtracted from X leaving the residual matrix E_1 (See Fig 13). The second PC is then fitted to the residuals E_1 , giving t_2 and p_2 . The procedure is repeated until all systematic variation has been captured.

The diagram illustrates the generative model structure. It shows a large box labeled X on the left, followed by an equals sign. To the right of the equals sign is a small vertical box labeled t , followed by a large box labeled X_{model} , followed by a plus sign, and finally a large box labeled E . Above the X_{model} box is a smaller box labeled p , with a downward arrow pointing from p to X_{model} .

Fig. 13. A presentation of the matrices involved in PCA. X is the spectral data from which the scores (t) and loadings (p) are calculated. E is the residual matrix containing the information not included in X_{model} .

A projection to latent structure-discriminant analysis (PLS-DA) is a linear regression method used to find the relationship between two datasets (X and y) in order to classify new objects. PLS-DA shares some similarities with PCA but the most important difference is the introduced y -vector, containing designed dummy variables that guide the algorithm to capture the relevant variation in the X matrix. A PLS can be explained simply as two interdependent PCA (one for X and one for y) connected via the score vectors. Although, the model is somewhat more complex than a PCA. It is important to understand that the captured variance in a PLS-DA model can be quite different compared to the captured variance in a PCA, this is due to the supervised guidance of the X matrix via the dummy variables in the y -vector.

When a multivariate model is fitted to a set of data the amount and relevance of the captured variance (PCA and PLS-DA), and ability to predict new samples, (PLS-DA) must be validated so that the optimal numbers of PCs can be decided [46, 48]. There are multiple ways of assessing this and three methods mainly used in this study are described below. Firstly, the residuals of X (E in Fig 13.) reveal if the systematic variation has been captured. The PCs are ordered according to described amount of variance, the first few PCs will describe the underlying phenomenon [49] that is searched for and the latter ones describe only noise. Thus one can study the amount of described variance and decide number of components based on the break point between systematic variation and noise. This, however, usually leads to an excessive number of included PCs because not even all the systematic variation might be of interest in a PLS-DA. Another approach is to use validation techniques such as leave-one-out cross-validation (LOO-CV) and test-set validation. The LOO-CV is helpful when a limited number of objects are available. The basic principle of LOO-CV is to split the dataset, object-wise, and one object (row containing spectral data) is taken out at a time whilst a model is fitted with one to n number of latent variables, then the object left out is predicted. This is repeated until all objects have been excluded. Predicted residual error sum of squares ($PRESS$) can then be calculated for each number of included latent variables according to Eq. 1 and plotted against the corresponding number of latent variable.

$$PRESS = \sum_{i=1}^n (y_{\text{pred}} - y_{\text{obs}})^2 \quad \text{Eq. 1}$$

By this procedure the appropriate model can be decided. Furthermore, a test-set is used to assure the appropriate number of latent variables. The squared sum of error of prediction (Q^2) for the test-set can then be compared for each number of included latent variables. Typically, when comparing the three methods for assessment, the optimum number of latent variables decreases when going from investigating the residuals to LOO-CV and test-set for assessment.

R^2 , Q^2 , and captured X -variation are important statistical aspects of a model. R^2 (Eq. 2) is the modeled (included) variance of the y -vector compared to the total variance of the y -vector whereas Q^2 (Eq. 3) is the predicted variance of an unknown or cross-validated sample compared to the total variance of the predicted sample. The captured X -variance is the amount of variance included in the model compared to the total variance of the spectral data.

$$R^2 = 1 - \sum_{i=1}^n (y_{\text{calc}} - y_{\text{obs}})^2 / \sum_{i=1}^n (y_{\text{obs}} - y_{\text{mean}})^2 \quad \text{Eq. 2}$$

$$Q^2 = 1 - PRESS / \sum_{i=1}^n (y_{\text{obs}} - y_{\text{mean}})^2 \quad \text{Eq. 3}$$

The purpose of PLS-DA is classification of unknown objects. A new y -value is calculated when the measured spectrum of an unknown object (compound) is predicted. In order to arrange the predicted y -values as members or non-members class intervals must be set. Decision limits can be decided in several ways, in this study an interval was based on the pooled variance of predictions according to Eq. 4 where S_1 is the standard deviation and n_1 the number of objects in member class and the prediction (S_{-1} and n_{-1} corresponds to non-members).

$$S_{pooled} = \left\{ \frac{[(n_1 - 1)S_1^2 + (n_{-1} - 1)S_{-1}^2]}{(n_1 + n_{-1} - 2)} \right\}^{1/2} \quad \text{Eq. 4}$$

The interval was then set to $1.0 \pm 3S_{pooled}$ for membership and $-1.0 \pm 3S_{pooled}$ for non-membership.

Soft independent modeling of class analogy (SIMCA) was applied to investigate the relations between compounds **7**, **8**, **9**, and **10** with respect to the pentose R⁴-substituent. Wold *et al.* first introduced SIMCA in the 1970s, as a pattern recognition technique [50]. SIMCA is a modelling technique applied to objects related to soft classes. When two classes overlap they are considered to be soft. The actual classification technique is not applied in this investigation but the modelling power (M_i , Eq. 5) and discriminatory power (D_i , Eq. 6), related to SIMCA, were calculated in order to scrutinise the X matrix by PCA.

$$M_i = 1 - S_{iresidual} / S_{iraw} \quad \text{Eq. 5}$$

Modelling power is calculated in Eq. 5, where S_{iraw} is the standard deviation of the variable in the original spectral data and $S_{iresidual}$ is the standard deviation of the variable in the residuals in E . The obtained values for each variable varies between one and zero, where variables with M_i close to zero are of little or no use for describing the underlying phenomenon.

$$D_i = \sqrt{\frac{\text{classA modelB } S_{iresidual}^2 + \text{classB modelA } S_{iresidual}^2}{\text{classA modelA } S_{iresidual}^2 + \text{classB modelB } S_{iresidual}^2}} \quad \text{Eq. 6}$$

Discriminatory power (D_i , Eq. 6) reveals the variables that discriminate between two classes. The objects (spectral data corresponding to each compounds) must be split into two classes before D_i can be calculated, the splitting must reflect the differences (substituents in this investigation). Thus **7** and **8** formed the first class (A) and **9** and **10** the second (B). PCA is then carried out providing two models, one for each class. Thereafter all objects in the first class are fitted to both models and the squared standard deviation of each variable in E is inserted in Eq. 6 and likewise for the second class (B). Then D_i is calculated and a larger value indicates a higher discriminatory power, the variables (spectral data points) holding the largest D_i are most valuable to the discrimination between classes.

Classification of saponins by multivariate modelling

The mass spectra obtained of isolated compounds of fractions QH-A, QH-B and QH-C were imported into Matlab™, where all calculations were performed. The spectra have several sources of variation due to noise, concentration differences, fragmentation patterns, stability and molecular weight of the precursor ion. The aim of the developed method is to identify saponins with respect to fragmentation pattern. Scaling was used to make spectral data comparable with respect to noise, concentration and ion abundance whereas the impact of molecular weight were to be levelled out. In paper I (QH-A) data was normalized whilst in paper II (QH-B and QH-C) standard normal variate correction (SNV) [48] was applied. Spectra were re-referenced in order to remove the impact of molecular weight and PCA was calculated followed by the classification models (PLS-DA). The y vectors were constructed before PLS-DA could be initialised, a total of 9 y vectors were constructed for QH-A and 12 for QH-B and QH-C. Each of the variables code for a structural element at a certain position. Furthermore, the vital part of this method is the re-referencing of spectral data so that the $[M+Na]^+$ is the new reference on the m/z axis. The constructed spectra correlate the peaks to the fragmentation of the saponins and remove the influence of molecular weight. The impact of referencing is visualised on spectral data from compounds in QH-A in Fig. 14 by two PCA models of the same MS^2 data before (a) and after (b) re-referencing. The score plot with referenced data in Fig. 14(b) has the saponins grouped due to loss of fragments in R^1 whereas the groups in Fig. 14(a) are uncorrelated to structural information in R^1 .

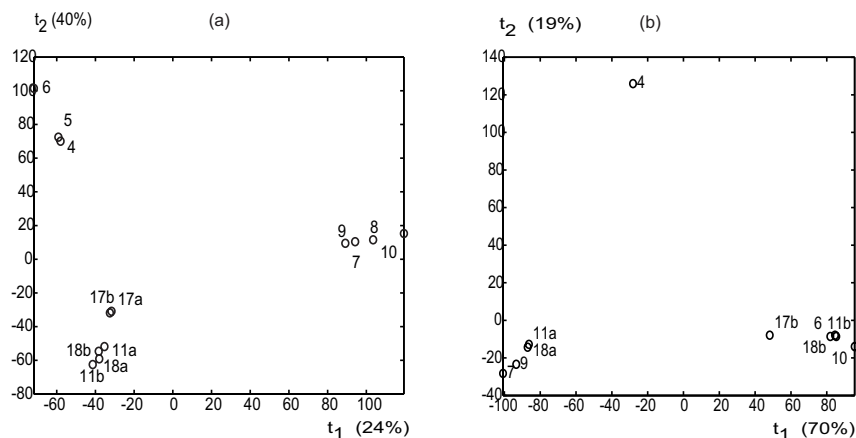


Fig. 14. Two PCA models of QH-A saponins. The score plots show MS^2 data (a) before re-referencing, (b) after re-referencing, data was normalised before a PCA was performed.

Fig. 15 shows the successive transformation of MS^3 spectral data of compound **6** via re-referencing and normalisation to the loading plots (PLS-DA) and the connection to the fragmentation route. The precursor ion $[B+Na]^+$ was isolated

and fragmented. The fragments $[C+Na]^+$, $[D+Na]^+$, $[E+Na]^+$, and $[S+Na]^+$ (See Fig. 11) are all pinpointed throughout the procedure. $[S+Na]^+$ is an end charged element that is separated from the fucose unit whereas the other fragments are based on a fucosyl residue ion.

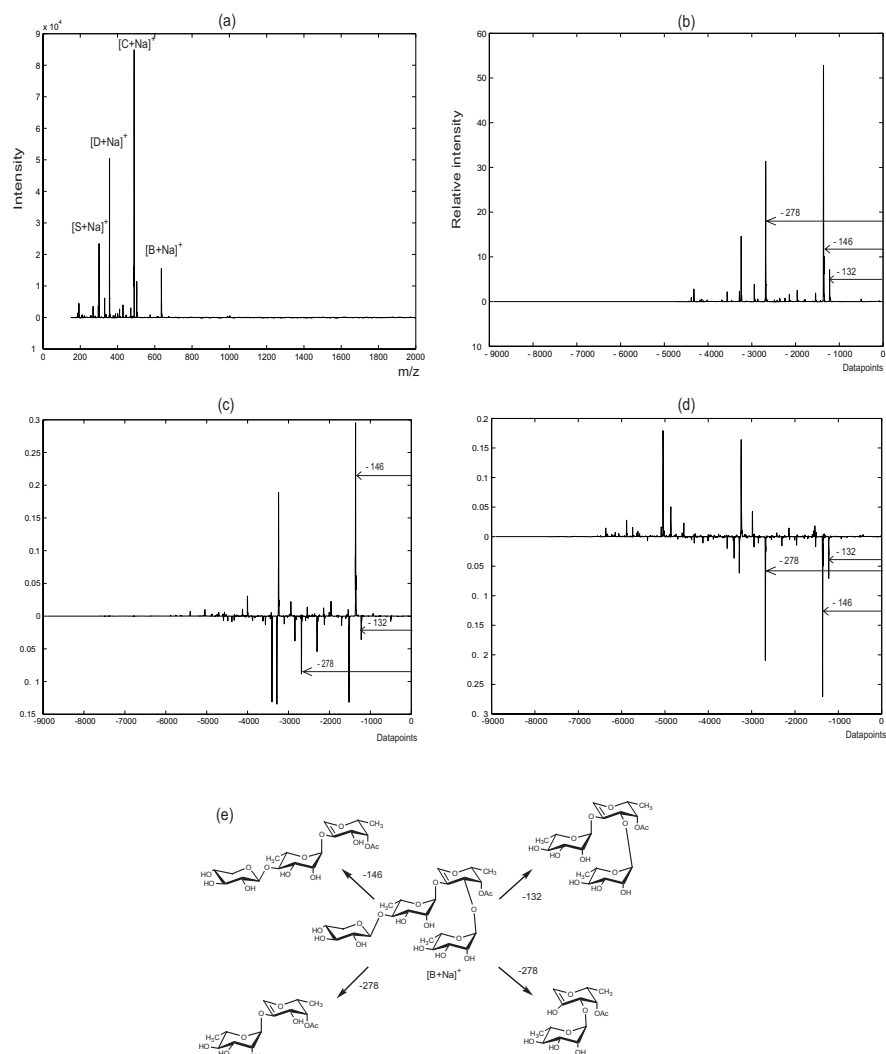


Fig. 15 (a) MS^2 spectra of compound **6** with proposed fragments, (b) Processed spectra. (c) Loading vector 1 and (d) loading vector 2 from a PLS-DA model. (e) Proposed fragmentation of precursor ion $[B+Na]^+$ at m/z 635.2 from compound **6**. Loss of key fragment masses are given in *italics*.

The obtained models were then validated and the optimal number of PCs were decided with respect to $PRESS$, Q^2_{cv} and $Q^2_{test-set}$. The number of PCs used was 2 or 3 for QH-A and two and seven for QH-B and QH-C models. In paper I the decision limits were based on the cross-validation, in order to assess the

predictions of the test-set. In paper II decision limits were based on the test-set predictions in order to evaluate the spectral data obtained by LC/MS. Test-set validation is anticipated to simulate the practical application of a model on future spectral data hence the calculations of decision limits based on the test-set is the best estimate.

Differentiation of apiose from xylose as R⁴-substituent.

In all three investigated fractions of saponins there are positions in the structure containing a pentose unit that is either xylose or apiose. The PLS-DA models described above proved to be insufficient for distinguishing these two closely related substituents. There are four compounds in the QH-A fraction with the actual pentose substituent, compounds **7** and **8** have an apiose and compounds **9** and **10** xylose as R⁴-substituent. The fragmentation amplitude was ramped between 0.4 V and 0.7 V in 0.03 V increments, resulting in spectral data from in total 44 MS³ spectra of each compound. Spectral data was initially evaluated using the nine PLS-DA classification models (paper I), resulting in 25 positively identified objects of which 23 were used in this study. The two excluded but positively identified objects were recognized as outliers in the initial PCA, they corresponded to the lowest identified fragmentation amplitude. The exploratory analysis of the 23 re-referenced, normalized and mean centered spectra (Fig. 16(a)) using PCA showed a promising result with two groups corresponding well to the R⁴-substituents (Fig. 16(b)).

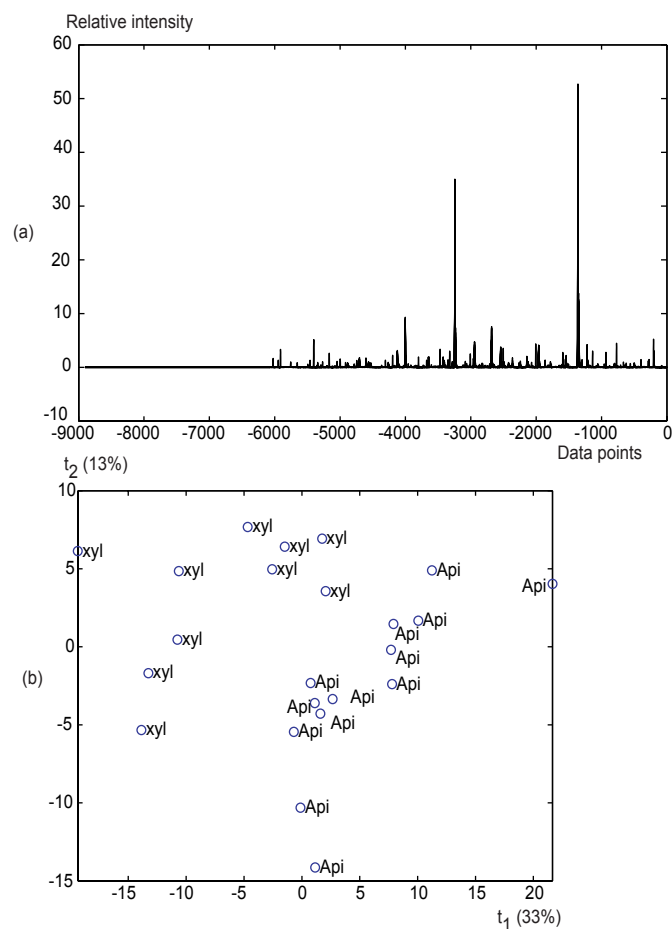


Fig. 16. (a) Shows the pretreated 23×8901 data points corresponding to the selected spectra. (b) Shows the corresponding PCA of the 23 objects.

Modelling power and discriminatory power, related to SIMCA, were used to outline the important variables of the underlying phenomena analogous to the classification of these objects. Usually, utilizing SIMCA, the objects are first divided class-wise and then modelling and discriminatory power are calculated. In this investigation the class-wise data-split was outlined only for the discriminatory power calculation. The motive is that expectedly the same variables (data points) are important to the PCA modelling in both classes since the fragmentation pattern is the same, thus splitting is not beneficial. The difference between the classes containing apiose and xylose is expected in relative intensities. The modelling power was calculated for all 23 objects as one class according to Eq. 5, 1811 variables (data points) were found to be the most contributing variables. The objects were then divided class-wise with **7** and **8** (13 objects with apiose as R^4 -substituent) in one class and **9** and **10** (10 objects with xylose as R^4 -substituent) in the other class. The variables with high modelling power were then selected in each class thus yielding a 13×1811 X -matrix and a

10×1811 X -matrix. Discriminatory power was then calculated according to Eq. 6 for each of the 1811 variables. Fig 17(a) shows the area of the most important variables for discrimination between classes, the peaks assigned **I** and **II** correspond to the main fragment-ions in Fig 16(a) above. Fig 17(b) shows the score plot from the PCA calculated on the corresponding 23×112 matrix.

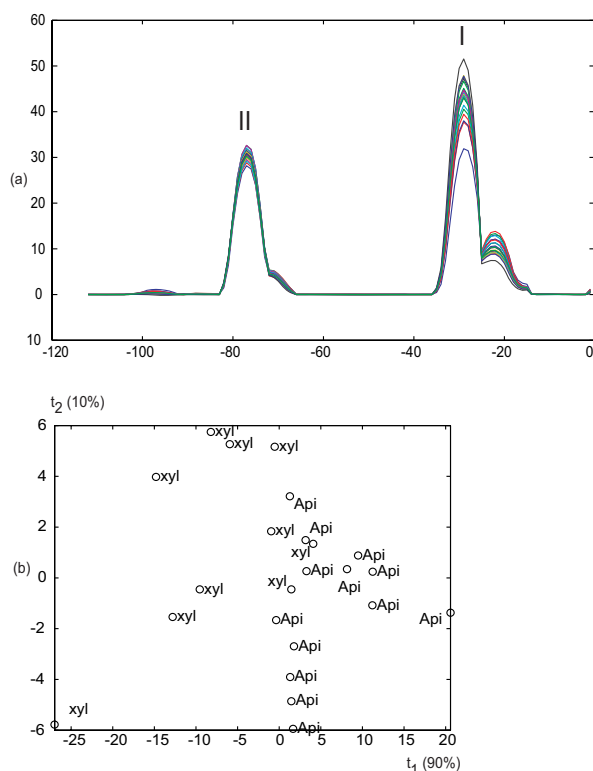


Fig. 17. (a) Shows the 112 selected data points corresponding to the 23 selected spectra. (b) Shows the corresponding PCA of the 23×112 matrix.

The peak in Fig. 17(a) assigned as **I** corresponds to the fragment-ion $[\mathbf{C}+\text{Na}]^+$ and the peak assigned as **II** corresponds to the fragment-ion $[\mathbf{S}+\text{Na}]^+$. The PCA loading plot that corresponds to the score plot in Fig 17(b) suggests that the fragment-ion $[\mathbf{C}+\text{Na}]^+$ assigned as **I** in Fig. 17(a) has a higher relative intensity with apiose as R⁴-substituent. The fragmentation route of $[\mathbf{B}+\text{Na}]^+$ producing the fragment ions $[\mathbf{C}+\text{Na}]^+$ and $[\mathbf{S}+\text{Na}]^+$ is visualized in Fig. 18.

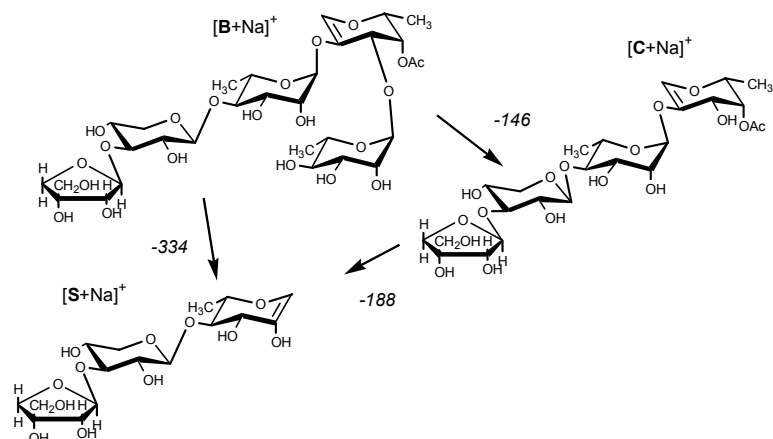


Fig 18. The fragment ions important to differentiate between the xylose and apiose substituents in compound 7, 8 and 9, 10.

Finally, a PLS-DA was fitted to the spectral data consisting of an X matrix of 23×8901 variables, and a y vector where +1 corresponded to objects containing a xylose and -1 to objects containing an apiose as R^4 -substituent. PRESS reached its first minimum with three latent variables, and R^2 and Q^2_{cv} were calculated to 0.97 and 0.82, respectively. The score plots consisting of the three latent variables plotted in Fig. 19 show the 10 xylose objects as one group and the 13 apiose objects as another group.

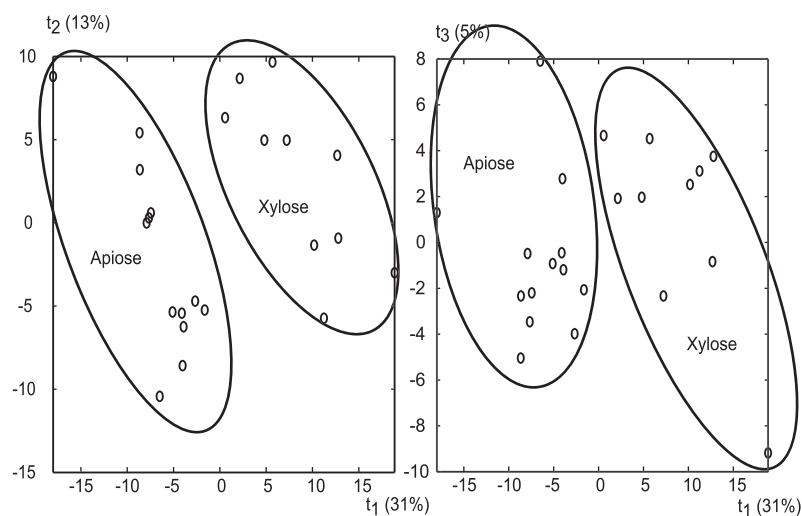


Fig. 19. Final PLS-DA score plots of 23 objects corresponding to selected spectra of compounds 7, 8, 9 and 10. Two groups with different structural features of the R^4 -substituent, xylose and apiose.

Classification of previously unreported saponins

A fraction of QH-C (Fig. 4(d)) containing saponins previously not reported was investigated in paper II. The spectral data corresponding to MS² and MS³ were extracted and pretreated. The 12 PLS-DA models were then used for prediction of y-values analogous to each substituent. Table 3 shows the result of the predictions.

Table 03. Predictions for unknown objects according to 12 structural features.

Compound	PLS-DA model					
	R ¹ , H	R ¹ , Xyl	R ¹ , Rha	R ² , H	R ² , pentose	R ³ , Fa-Ara
J1	0.9	-0.9	-0.9	-0.8	0.8	0.6
J1a	0.9	-1.0	-0.9	-1.0	1.0	-0.8
J2	-1.0	-1.0	1.0	1.0	-1.0	-1.8
J3	-1.0	1.0	-1.0	0.7	-0.7	-1.5
	R ³ , H	R ³ , Ac	R ⁴ , Fa-Ara	R ⁴ , H	R ⁵ , Glc	R ⁵ , H
J1	-0.6	-0.9	-0.6	0.6	-0.8	0.8
J1a	0.8	-1.0	0.8	-0.8	-0.8	0.8
J2	1.0	-0.2	1.8	-1.8	0.3	-0.3
J3	0.7	-0.1	1.5	-1.5	0.1	-0.1

^a Bold-face values indicate non-classified objects.

The compounds **J1** and **J1a** were classified and a structure was proposed, see Table 4. The molecular weight of both **J1** and **J1a** were calculated to 1858.0 that correspond well to the proposed structure. **J2** and **J3** were predicted outside the decision limits indicating structural features unknown to the PLS-DA models.

Table 04. Proposed substituents in compound **J1** and **J1a**. Numbering of substituents correspond to Fig. 3.

Compound	R ¹	R ²	R ³	R ⁴	R ⁵
J1	H	Pentose (apiose or xylose)	Fa-Ara	H	H
J1a	H	Pentose (apiose or xylose)	H	Fa-Ara	H

LC/MS on chromatographic fractions containing saponins

The PLS-DA methods outlined was applied to LC/MS data, in order to find out whether classification of such data is possible. The recording of a spectrum showed to be a limitation, data obtained from HPLC on-line coupled electrospray mass spectrometry has a much narrower time window in which the acquiring can take place compared to direct injection using a syringe pump. Direct injection was used to gather data for PLS-DA modelling. When data was acquired using direct injection the accumulation time of the trap varied between 0.1-50 ms, and up to 512 scans were averaged to produce the final spectra. The resulting time window needed for this measurement is up to several minutes. The time during which a compound is possible to measure using HPLC in the

presented method is shorter, less than 2 minutes. As a result the accumulation time was limited to between 0.1 and 2 ms and 64 scans were averaged. Three injections were performed in order to acquire spectra in MS¹, MS², and MS³. Two different previously separated fractions (phosphate buffer at pH 6.4) were investigated as well as QH-B and QH-C. The aim of the pre-separated fractions was to prove classification of regio-isomers possible in LC/MS data whereas the reason for investigating QH-B and QH-C was to show the appropriateness of the method applied on more complex fractions. The two pre-separated fractions contained the compounds **S1**, **S1a**, **S2**, **S2a** and **B1**, **B1a**, **B2**, and **B2a**, respectively. The fractions mainly contain pairs of saponins with differences in R¹ (14 Da, xylose, rhamnose) and with both fatty acyl regio isomers visualized in Fig. 4(a) and Fig. 4(h). The classification of the regio-isomers showed to be possible; Table 5 shows predicted values of classes corresponding to MS³ data. **S1**, **S2**, **B1**, **B2** were studied in the two chromatographic fractions QH-B and QH-C. The obtained data was pre-treated and predicted using the PLS-DA models in paper II, the result can be seen in Table 6. The positive identification of compounds by multivariate analysis of HPLC on-line coupled ES spectral data has shown that the method is valid for use on mixtures of saponins.

Table 05. Prediction of MS^3 spectral data from LC/MS on pre-separated fractions. A negative value indicates a non-member and a positive a member of the class. Each class represents an R-substituent.

Compound	PLS-DA model								
	R^2 , H	R^2 , pentose	R^3 , Fa-Ara	R^3 , H	R^3 , Ac	R^4 , Fa_Ara	R^4 , H	R^5 , Glc	R^5 , H
S1	0.9	-0.9	-0.8	0.8	-1.0	0.9	-0.9	-0.7	0.7
S1a	0.8	-0.8	0.6	-0.7	-1.0	-0.6	0.6	-0.8	0.8
S2	0.9	-0.9	-0.8	0.7	-0.8	0.9	-0.9	-0.7	0.7
S2a	1.1	-1.1	0.7	-0.9	-0.8	-0.8	0.8	-0.7	0.7
B1	0.7	-0.7	-0.4^a	0.3^a	-0.9	0.4^a	-0.4^a	0.6	-0.6
B1a	0.4	-0.4	0.8	-0.8	-1.0	-0.8	0.8	0.5^a	-0.5
B2	1.1	-1.1	-0.9	0.8	-0.9	0.9	-0.9	0.7	-0.7
B2a	0.4	-0.4	0.7	-0.6	-1.1	-0.7	0.7	0.6	-0.6

^a Bold-face values indicate non-classified objects.

Table 06. Prediction of peaks from LC/MS on QH-B and QH-C. A negative value indicates a non-member and a positive a member of the class. Each class represents R-substituent.

Compound	PLS-DA model					
	R^1 , H	R^1 , Rha	R^1 , Xyl	R^2 , H	R^2 , pentose	R^3 , Fa-Ara
B1	-0.9	-0.8	0.7^a	0.7	-0.7	-1.1
B2	-0.8^a	0.7^a	-0.9	0.9	-0.9	-0.7
S1	-0.9	-0.9	0.8	0.9	-0.9	-0.7
S2	-1.0	0.9	-1.0	0.8	-0.8	-1.1
	R^3 , H	R^3 , Ac	R^4 , H	R^4 , Fa_Ara	R^5 , H	R^5 , Glc
B1	1.0	-0.9	-1.1	1.1	-1.1	1.1
B2	0.8	-1.1	-0.7	0.7	-1.0	1.0
S1	0.9	-1.2	-0.8	0.8	0.8	-0.8
S2	1.1	-1.1	-1.1	1.1	0.9	-0.9

^a Bold-face values indicate non-classified objects.

Concluding remarks

Today the most promising formulation of saponins as adjuvants is ISCOM[®]s. The way these microspheres interact in humans are not fully understood, neither is the function of saponins in human. However, a mixture of *Quillaja* saponins from QH-A and QH-C with low toxicity and high adjuvant activity has been proposed that are being tested in clinical trials.

The method outlined here is an analytical tool for further understanding and applications of *Quillaja* saponins. Most of the major compounds in saponin fractions from *Quillaja saponaria* Molina have been structurally elucidated by NMR studies, which is not the case for the minor compounds that are still to be investigated. Applying NMR techniques usually require around 200 µg or more of the isolated sample, compared to a few µg for a full classification using mass spectrometry. Combining mass spectrometry with multivariate analysis provides a rapid method for structural classification of these minor unknown compounds. Further, young plants of *Quillaja saponaria* Molina have shown different metabolomic profiles of saponins compared to older specimen; hence the biological and chemical activity differs between batches from different specimen. The LC/MS method outlined in paper II could in future work be used together with multi-way analysis for screening of chromatographic fractions of saponins or to monitor the metabolomics of *Quillaja saponaria* Molina saponins, in order to achieve better and more identical fractions for use as adjuvants as well as to study variations between batches and species.

The method developed in paper I and II is specific to a group of *Quillaja* saponins, but can be applied to other molecules built up of monomers such as glycoproteins.

References

1. San Martín, R. and R. Briones, *Industrial Uses and Sustainable Supply of Quillaja saponaria (Rosaceae) saponins*. Economic Botany, 1999. **53**(3): p. 302-311.
2. Dalsgaard, K., *Thin layer chromatographic fingerprinting of commercially available saponins*. Dansk Tidsskrift for Farmaci, 1970. **44**: p. 327-331.
3. Strobbe, R., et al., *Studies about the adjuvant activity of saponin fractions in foot and mouth disease vaccine. III. Comparison of the irritant, adjuvant and hemolytic activities of six commercial saponins and their hemolytic fractions obtained by chromatography on sephadex G 100*. Arch Exp Veterinarmed, 1976. **30**(2): p. 173-81.
4. Kensil, C.R., et al., *Separation and characterization of saponins with adjuvant activity from Quillaja saponaria Molina cortex*. The Journal of Immunology, 1991. **146**(2): p. 431-437.
5. Kensil, C.R., *Saponins as Vaccine Adjuvants*. Critical Reviews in Therapeutic Drug Carrier Systems, 1996. **13**(1-2): p. 1-55.
6. Dalsgaard, *A study of the isolation and characterization of the saponin Quil A. Evaluation of its adjuvant activity, with a special reference to the application in the vaccination of cattle against foot-and-mouth disease*. Acta Vet. Scand. Suppl., 1978(69): p. 7-40.
7. Kersten, G.F., et al., *Immunogenicity of liposomes and iscoms containing the major outer membrane protein of Neisseria gonorrhoeae: influence of protein content and liposomal bilayer composition*. Infect Immun, 1988. **56**(6): p. 1661-4.
8. Morein, B., et al., *Martix with immunomodulating activity*. 1989.
9. Rönnberg, B., M. Fekadu, and B. Morein, *Adjuvant activity of non-toxic Quillaja saponaria Molina components for use in ISCOM matrix*. Vaccine, 1995. **13**(14): p. 1375-82.
10. Kensil, C.A., S. Soltysik, and D. Marciani, *Saponin-antigen conjugates and the use thereof*. 1993.
11. Aucouturier, J., L. Dupuis, and V. Ganne, *Adjuvants designed for veterinary and human vaccines*. Vaccine, 2001. **19**: p. 2666-2672.
12. Morein, B., et al., *Iscom, a novel structure for antigenic presentation of membrane proteins from enveloped viruses*. Nature, 1984. **308**(5958): p. 457-60.
13. Sjölander, A., et al., *Immune responses to ISCOM formulations in animal and primate models*. Vaccine, 2001. **19**(17-19): p. 2661-5.
14. Barr, I.G., A. Sjölander, and J.C. Cox, *ISCOMs and other saponin based adjuvants*. Advanced Drug Delivery Reviews, 1998. **32**(3): p. 247-271.
15. Sanders, M.T., et al., *ISCOM-based vaccines: The second decade*. Immunology and Cell Biology, 2005. **83**: p. 119-128.
16. Higuchi, R., et al., *Structure of desacylsaponins obtained from the bark of Quillaja saponaria*. Phytochemistry, 1987. **26**(1): p. 229-235.
17. Guo, S.J., et al., *Triterpenoid saponins from Quillaja saponaria*. Phytochemistry, 1998. **48**(1): p. 175-180.
18. Guo, S.J., et al., *Triterpenoid saponins containing an acetylated branched D-fucosyl residue from Quillaja saponaria Molina*. Phytochemistry, 2000. **53**(8): p. 861-868.
19. Guo, S.J. and L. Kenne, *Characterization of some O-acetylated saponins from Quillaja saponaria Molina*. Phytochemistry, 2000. **54**(6): p. 615-623.
20. Guo, S.J. and L. Kenne, *Structural studies of triterpenoid saponins with new acyl components from Quillaja saponaria Molina*. Phytochemistry, 2000. **55**(5): p. 419-428.

21. Nord, L.I. and L. Kenne, *Novel acetylated triterpenoid saponins in a chromatographic fraction from Quillaja saponaria Molina*. Carbohydrate Research, 2000. **329**(4): p. 817-829.
22. Nord, L.I. and L. Kenne, *Separation and structural analysis of saponins in a bark extract from Quillaja saponaria Molina*. Carbohydrate Research, 1999. **320**(1-2): p. 70-81.
23. Nyberg, N.T., et al., *Separation and structural analysis of some saponins from Quillaja saponaria Molina*. Carbohydrate Research, 2000. **323**: p. 87-97.
24. Jay, A., *The methylation reaction in carbohydrate analysis*. Journal of Carbohydrate Chemistry, 1996. **15**: p. 897-923.
25. van Setten, D.C., et al., *Glycosyl Compositions and Structural Characteristics of the Potential Immuno-adjuvant Active Saponins in the Quillaja saponaria Molina Extract Quil A*. Rapid Communications in Mass Spectrometry, 1995. **9**: p. 660-666.
26. van Setten, D.C., et al., *Multiple-Stage Tandem Mass Spectrometry for Structural Characterization of Saponins*. Analytical Chemistry, 1998. **70**(20): p. 4401-4409.
27. van Setten, D.C., et al., *Ion Trap Multiple-Stage Tandem Mass Spectrometry as a Pre-NMR Tool in the Structure Elucidation of Saponins*. Phytochemical Analysis, 2000. **11**(3): p. 190-198.
28. Nord, L.I., L. Kenne, and S.P. Jacobsson, *Multivariate analysis of ¹H NMR spectra for saponins from Quillaja saponaria Molina*. Analytica Chimica Acta, 2001. **446**(1-2): p. 199-209.
29. Chen, M., et al., *Analysis of flavonoid constituents from leaves of Acanthopanax senticosus harms by electrospray tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2002. **16**(4): p. 264-71.
30. Broberg, S., L.I. Nord, and L. Kenne, *Oligosaccharide sequences in Quillaja saponins by electrospray ionization ion trap multiple-stage mass spectrometry*. Journal of Mass Spectrometry, 2004. **39**(6): p. 691-701.
31. Broberg, S., *Studies of Oligo- and Polysaccharides by MALDI-TOF and ESI-ITMSⁿ Mass Spectrometry*, in Organic chemistry. 2004, Swedish University of Agricultural Sciences: Uppsala.
32. Huhman, D.V. and L.W. Sumner, *Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer*. Phytochemistry, 2002. **59**(3): p. 347-360.
33. Kite, G.C., M.J.R. Howes, and M.S.J. Simmonds, *Metabolomic analysis of saponins in crude extracts of Quillaja saponaria by liquid chromatography/mass spectrometry for product authentication*. Rapid Communications in Mass Spectrometry, 2004. **18**(23): p. 2859-2870.
34. Cui, M., et al., *Rapid identification of saponins in plant extracts by electrospray ionization multi-stage tandem mass spectrometry and liquid chromatography/tandem mass spectrometry*. Rapid Communications in Mass Spectrometry, 2000. **14**(14): p. 1280-1286.
35. Bin, L., et al., *Rapid structural characterisation of triterpenoid saponins in crude extract from Symplocos chinensis using liquid chromatography combined with electrospray ionisation tandem mass spectrometry*. Journal of Chromatography A, 2006. **1101**: p. 53-62.
36. Li, R., et al., *ESI-QqTOF-MS/MS and APCI-IT-MS/MS analysis of steroid saponins from the rhizomes of Dioscorea panthaica*. Journal of Mass Spectrometry, 2006. **41**: p. 1-22.
37. Harvey, D.J., *Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates*. Mass Spectrometry Reviews, 1999. **18**: p. 349-451.

38. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionisation fundamentals*. Mass Spectrometry Reviews, 2001. **20**: p. 362-387.
39. Gaskell, S.J., *Electrospray: Principles and Practice*. Journal of Mass Spectrometry, 1997. **32**: p. 677-688.
40. Kebarle, P., *A brief overview of the present status of the mechanism involved in electrospray mass spectrometry*. Journal of Mass Spectrometry, 2000. **35**: p. 804-817.
41. March, R.E., *An Introduction to Quadrupole Ion Trap Mass Spectrometry*. Journal of Mass Spectrometry, 1997. **32**: p. 351-369.
42. Box, G.E.P., W.G. Hunter, and J.S. Hunter, *Statistics for experimenters*. 1978, New York: John Wiley & Sons.
43. Riter, L.S., et al., *Statistical design of experiments as a tool in mass spectrometry*. Journal of Mass Spectrometry, 2005. **40**(5): p. 565-579.
44. Brereton, R., *Chemometrics "Data analysis for the Laboratory and Chemical Plant"*. 2003, Chichester: John Wiley & Sons.
45. Smilde, A., R. Bro, and P. Geladi, *Multi-way Analysis "Applications in the chemical sciences"*. 2004, Chichester: John Wiley & Sons.
46. Wold, S., K. Esbensen, and P. Geladi, *Principal Component Analysis*. Chemometrics and Intelligent Laboratory Systems, 1987. **2**(1-3): p. 37-52.
47. Geladi, P. and B.R. Kowalski, *Partial Least-Squares Regression - a Tutorial*. Analytica Chimica Acta, 1986. **185**: p. 1-17.
48. Eriksson, L., et al., *Multi- and Megavariate Data Analysis*. 2nd ed. Vol. 1. 2006, Umeå: Umetrics academy.
49. Geladi, P., M. Manley, and T. Lestander, *Scatter plotting in multivariate data analysis*. Journal of Chemometrics, 2003. **17**: p. 503-511.
50. Wold, S. and M. Sjöström. *SIMCA, a method for analyzing chemical data in terms of similarity and analogy*. in *Chemometrics: theory and application*. 1976. San Francisco, Calif.

Acknowledgments

I would like to thank my supervisor Professor Lennart Kenne for introducing me in general to science and in particular to carbohydrate chemistry. You have an exceptionally trained eye when it comes to science and I am very grateful to have this knowledge within my grasp. Your attitude and humour are of course always welcome around at the department and I appreciate it greatly.

I would also like thank Lars Nord, my assistant supervisor and former co-worker, for further introducing me to chemometrics and for still being supportive to my work.

Next, I am much thankful for all help and support from the people at the Department of Chemistry. There are also people I would like to direct my gratitude to:

Jan Eriksson and Suresh Gohil for many long and sometimes frustrating but also interesting moments in front of or behind the mass spectrometers.

Rolf Andersson and Corine Sandström for helping me with NMR problems, I am still sorry about the HR-MAS probe...not much are written about NMR in this thesis but hopefully will be in future studies.

I would also like to thank my family and my friends outside the lab for support.

Finally, this work was supported by grants from the Swedish Research Council.

Gratia